Lecture Module - Probability and Statistics

ME3023 - Measurements in Mechanical Systems

Mechanical Engineering

Tennessee Technological University

Module 4 - Probability and Statistics

Module 4 - Probability and Statistics

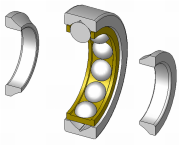
- Topic 1 - Histograms and Probability Density Functions
- Topic 2 - Characterizing a Population of Data
- Topic 3 - Z Table Examples

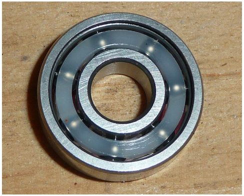
Topic 1 - Histograms and Probability Density Functions

- A Population of Data
- Randomly Distributed Data
- Frequency Histogram
- Probability Density Function

A Population of Data

Consider a manufacturer that makes 100000 of ball bearings in a set.





- How does the manufacturer ensure the quality of the product?
- How the does the seller communicate this to the buyer?
- How does the engineer use this information?

Images: Wikipedia

A Population of Data

The data set, or population, is generated by taking measurements of individuals chosen randomly from the entire population. **Sampling** refers to repeated measurements of the **measured variable** under fixed operating condition.

$$X = \{x_1, x_2, x_3, ..., x_i, ..., x_{n-1}, x_n\}$$

Randomly Distributed Data

Our discussion will assume that the values in the data set are **randomly distributed** about a mean value. It is important to consider what this means.

Frequency Histogram

"A histogram is an accurate representation of the distribution of numer- ical data. It is an estimate of the probability distribution of a continuous variable (CORAL) and was first introduced by Karl Pearson.[1] It differs from a bar graph, in the sense that a bar graph relates two variables, but a histogram relates only one." - Wikipedia

Frequency Histogram

Consider sampling the 0.5 inch diameter ball bearings with $\sigma = .001$.

0.5016	0.4991	0.4981	0.5003	0.4988	0.5007	0.4994	0.5000	0.4999	0.4999	0.4988
0.4997	0.4991	0.4986	0.5000	0.5018	0.4996	0.5010	0.4991	0.5011	0.5006	0.5006
0.5014	0.5001	0.4997	0.4994	0.5001	0.5009	0.5009	0.4992	0.4983	0.5004	0.5005
0.5004	0.4992	0.5009	0.5003	0.5018	0.5004	0.4996	0.5014	0.4992	0.4986	0.5002
0.4986	0.4992	0.5003	0.5002	0.4999	0.4998	0.5013	0.5009	0.5003	0.4986	0.4991
0.5030	0.4995	0.5012	0.5005	0.4992	0.4995	0.5005	0.5000	0.5005	0.4993	0.4987
0.4992	0.5018	0.4976	0.5003	0.4991	0.5007	0.5002	0.4980	0.5004	0.5000	0.4997
0.5002	0.5004	0.4980	0.5011	0.5003	0.5013	0.5000	0.5001	0.5003	0.5000	0.5013
0.5004	0.4988	0.5003	0.4999	0.5004	0.4996	0.4999	0.5000	0.4993	0.5003	0.4996
0.4992	0.5008	0.4997	0.5004	0.4998	0.4998	0.5006	0.5002	0.5010	0.5014	0.5019
0.5008	0.4997	0.5002	0.4993	0.4995	0.5007	0.4986	0.5009	0.4999	0.5002	0.5010
0.4996	0.4993	0.4982	0.4994	0.4993	0.4995	0.5002	0.5004	0.4999	0.5005	0.4993

Data: Generated in MATLAB, see probabilty_statistics_topic1.m

Frequency Histogram

Draw Historgram by hand

Frequency Histogram

Probability Density Function

Activity: Complete as an individual, discussion with peers is encouraged.
Write a MATLAB progam to do the following:

1. Generate a population of data that is randomly distributed about a central mean to represent a repeated measurement or physical property. One suggestion is to use the built-in function *normrnd()* . Choose the population size, mean value, and standard deviation.
2. Show the data in a labeled figure with item number on the x-axis and measured value on the y-axis.
3. Create a histogram using the population of data. One option is to use the MATLAB function *histogram()*. Choose an appropriate bin size for the histogram to represent the data. The goal is to show the central tendancy and random distribution of the population of data.
4. Show a continuous probability density function to represent the data. The function should have the same mean value and standard deviation as used above.

Deliverables: Upload your MATLAB code to the appropriate folder as a .m file *Concept Example: Sketch Histograms and Probabilty Density Functions*. Include your name, date, and topic in the program file.

Probability Density Function

Topic 2 - Characterizing a Population of Data

- A Population of Data
- Variance and Standard Deviation
- Using the Z Table
- Example: Using the Z Table

A Population of Data

Review from topic 1:

Consider a manufacturing facility that makes ball bearings.

- How does the manufacturer ensure the quality of the product?
- How the does the seller communicate this to the buyer?
- How does the Engineer use this information?

A Population of Data

Review from topic 1:

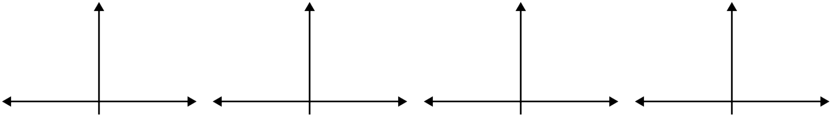
The data set, or population, is generated by taking measurements of individuals chosen randomly from the entire population. **Sampling** refers to repeated measurements of the **measured variable** under fixed operating condition.

$$X = \{x_1, x_2, x_3, ..., x_i, ..., x_{n-1}, x_n\}$$

A Population of Data

Consider the dataset as a population of data or a *distribution*.

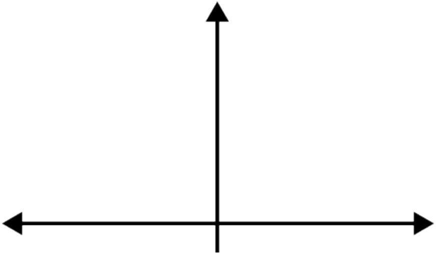
- What does the distribution look like?
- How can we describe it mathematically?
- What function p(x) appropriately describes the distribution?



A Population of Data

It has been proposed that physical systems behave randomly.

- population of data shows central tendancy about a mean value
- the distribution can be described by the Gaussian, or normal distribution
- the assumption of the distribution model can be used to make predictions about population



$$p(x) = \frac{1}{\sigma\,(2\pi)^{(1/2)}} exp\left[-\frac{1}{2}\frac{(x-x')^2}{\sigma^2}\right]$$

A Population of Data

For a given set of measurements we want to quantify:

- a representative value that characterizes the average of the measured data set
- a representative value that provides a measure of the variation in the data set
- how well the average of the measured data set represents the average of the entire population

Variance and Standard Deviation

Remember the histogram gernerated previously

$$N = \Sigma_{j=1}^{K} n_j$$

If you convert the value at each bin to a percentage, the area under the curve equals 100%.

$$100x\Sigma_{j=1}^{K} f_i = 100\%$$

Variance and Standard Deviation

Given the assumption that the variation in data is free from systematic errors, we can quantify the population with the following terms

$$x' = lim_{T\to\infty}\frac{1}{T}\int_0^T x(t)dt$$

$$x' = \int_{-\infty}^{\infty} xp(x)dx$$

For discrete data this becomes

$$x' = lim_{T\to\infty}\frac{1}{N}\Sigma_{i=1}^{N}x_i$$

Variance and Standard Deviation

The true variance is:

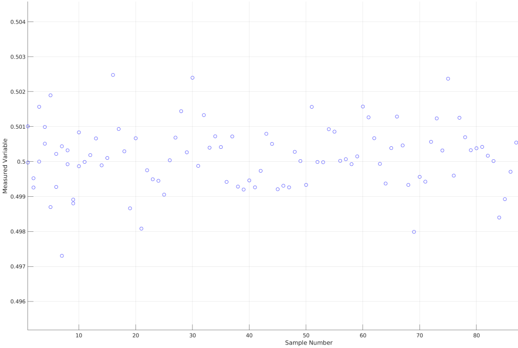
$$\sigma^2 = \int_{-\infty}^{\infty} (x - x')^2 p(x) dx$$

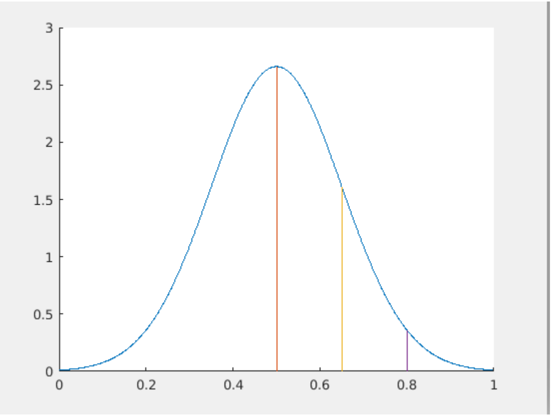
For discrete data this becomes:

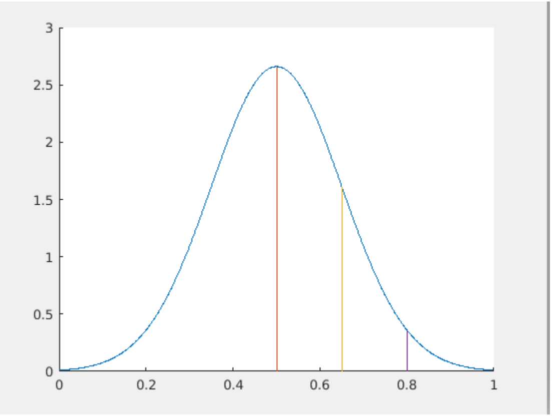
$$\sigma^2 = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} (x_i - x'$$

The square root of the **variance** is the **standard deviation**.

$$\sigma = \sqrt{\sigma^2}$$



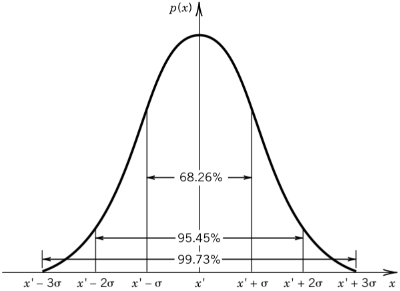

3
2.5
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1


3
2.5
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1

Using the Z Table

... The area under the portion of the probability density function curve, $p(x)$, defined by the interval $x' - z_1\sigma \leq x \leq x' + z_1\sigma$ provides the probability that a measurement will assume a value within that interval. Direct integration of $p(x)$ for a normal distribution between the limits $x' \pm z_1\sigma$ yields that for $z_1 = 1, 68.26\%$ of the area under $p(x)$ lies within $\pm 1\sigma$ x' . This means that there is a 68.26% chance that a measurement of x will have a value within the interval $x' \pm 1\sigma$.

Using the Z Table



$z_1 = 1, 68.26\%$ of the area under $p(x)$ lies within $\pm z_1\sigma$ of x'.
$z_1 = 2, 95.45\%$ of the area under $p(x)$ lies within $\pm z_1\sigma$ of x'.
$z_1 = 3, 99.73\%$ of the area under $p(x)$ lies within $\pm z_1\sigma$ of x'.

Using the Z Table

It follows directly that the representative value that characterizes a measure of the variation in a measured data set is the standard deviation. The probability that the ith measured value of x will have a value between $x' \pm z_1 \delta$ is $2 \times P(z_1) \times 100 = P\%$.
This is written as

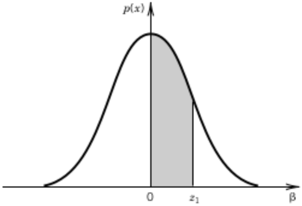
$$x_i = x' \pm z_1 \sigma \quad (P\%)$$

Using the Z Table

Example: Using the Z Table

Table 4.3 Probability Values for Normal Error Function: One-Sided Integral Solutions for $p(z_1) = \frac{1}{(2\pi)^{1/2}} \int_0^{z_1} e^{-\beta^2/2} d\beta$

$z_1 = \frac{x_1 - x'}{\sigma}$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1809	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2794	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4292	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4758	0.4761	0.4767
2.0	0.4772	0.4778	0.4803	0.4788	0.4793	0.4799	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.49865	0.4987	0.4987	0.4988	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990



Example: Using the Z Table

Topic 3 - Z Table Examples

- Table of Probability Values
- Example 1
- Example 2
- Example 3

Table of Probability Values

Table 4.3 Probability Values for Normal Error Function: One-Sided Integral Solutions for $p(z_1) = \frac{1}{(2\pi)^{1/2}} \int_0^{z_1} e^{-\beta^2/2} d\beta$

$z_1 = \frac{x_1 - x'}{\sigma}$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1809	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2794	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4292	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4758	0.4761	0.4767
2.0	0.4772	0.4778	0.4803	0.4788	0.4793	0.4799	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.49865	0.4987	0.4987	0.4988	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990

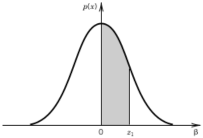


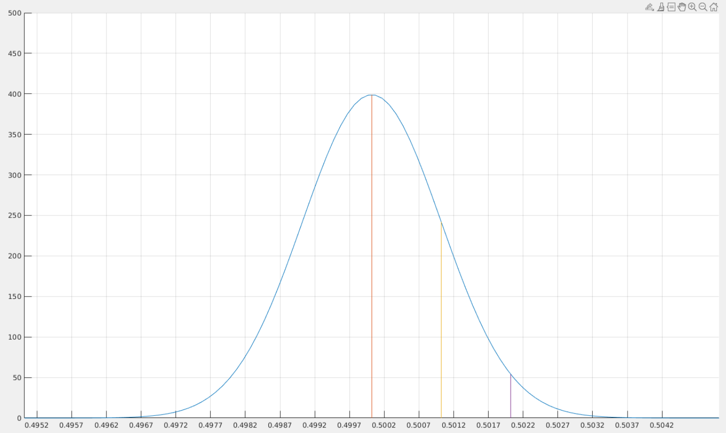
Table of Probability Values

Example 1

Using the probability values in Table 4.3, show that the probability that a measurement will yield a value within $x' \pm \sigma$ is 0.6826 or 68.26%.

Example 2

Now consider the measurements from the ball bearing factory in topic 1.



Example 2

If the true mean is 0.5 inches, what is the probability that a individual bearing will measure within 0.45 σ of the mean ? What about 0.15 σ?

What is the probability that a individual bearing will measure between 0.44 and 0.49 inches?

If the true mean is 0.5 inches, estimate the uncertainty interval at a probability of 70%. Also estimate the uncertainty interval at a probability of 90%. What is the probability that a individual bearing will measure within 0.45 σ of the mean ? What about 0.15 σ?

Example 3

What is the probability that a individual bearing will measure between 0.40 and 0.44 inches? What is different about this problem? What is the significance of the z value not present in the table?

Example 3